Weather

- **Step 1: Please create a decision tree for the following table.**

➢ 1st Iteration: Find the root of a decision tree

The **Parent Data Table** has Targets of **5N, 9Y** which produce entropy of **0.94**

Predictors				**Target**
Outlook	**Temp**	**Humidity**	**Windy**	**Play Golf**
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Overoact	Hot	High	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Overoact	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overoact	Mild	High	TRUE	Yes
Overoact	Hot	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No

5N,9Y

Prob(No) - 5/14 ==> 0.357142857

Prob(Yes) - 9/14 ==> 0.642857143

Entropy : -5/14 * log2(5/14) - 9/14 * log2(9/14) ==> 0.94

Information Gain for Outlook

Outlook	**Play Golf**
Rainy	No
Rainy	No
Rainy	No

Outlook	**Play Golf**	**Entropy**	**Prob(No)**	**Prob(Yes)**
Rainy	No	**3No,2Yes**	0.6	0.4
Rainy	No	**0.97**		
Rainy	No			
Rainy	Yes			
Rainy	Yes			

Rainy	Yes
Rainy	Yes
Overoact	Yes
Overoact	Yes
Overoact	Yes
Overoact	Yes
Sunny	Yes
Sunny	Yes
Sunny	Yes
Sunny	No
Sunny	No

Outlook	Play Golf	Entropy	Prob(Yes)
Overoact	Yes	**4 Yes**	1
Overoact	Yes	**0**	
Overoact	Yes		
Overoact	Yes		

Outlook	Play Golf	Entropy	Prob(No)	Prob(Yes)
Sunny	Yes	**3Yes,2 No**	0.4	0.6
Sunny	Yes	**0.97**		
Sunny	Yes			
Sunny	No			
Sunny	No			

Note:

The entropy of the **Information Gain for Outlook** -

0.94– (5/14 * 0.97 + 4/14 * 0 + 5/14 * 0.97) = **0.24714285714**

Note: 14 means 14 Outlook records

5 means 5 Rainy Outlook records

4 means 4 Overoact Outlook records

5 means 5 Sunny Outlook records

Information Gain for Temperature

Temp	Play Golf
Hot	No
Hot	No
Hot	Yes
Hot	Yes
Cool	Yes
Cool	No

Temp	Play Golf	Entropy	Prob(No)	Prob(Yes)
Hot	No	2No,2 Yes	0.5	0.5
Hot	No	**1**		
Hot	Yes			
Hot	Yes			

Temp	Play Golf	Entropy	Prob(No)	Prob(Yes)
Cool	Yes	1No,3 Yes	0.25	0.75
Cool	No	**0.81**		

Cool	Yes
Cool	Yes
Mild	Yes
Mild	Yes
Mild	Yes
Mild	No
Mild	No
Mild	Yes

Cool	Yes			
Cool	Yes			

Temp	Play Golf	Entropy	Prob(No)	Prob(Yes)
Mild	Yes	2 No,4 Yes	0.333333333	0.666666667
Mild	Yes	**0.92**		
Mild	Yes			
Mild	No			
Mild	No			
Mild	Yes			

Note:

The entropy of the **Information Gain for Temperature** -

0.94– (4/14 * 1 + 4/14 * 0.81 + 6/14 * 0.92) = 0.02857142857

Information Gain for Humidity

Humidity	Play Golf
High	No
High	No
High	Yes
High	Yes
High	No
High	Yes
High	No
Normal	Yes
Normal	Yes
Normal	Yes
Normal	Yes
Normal	Yes
Normal	No
Normal	Yes

Humidity	Play Golf	Entropy	Prob(No)	Prob(Yes)
High	No	4 No,3 Yes	0.571428571	0.428571429
High	No	**0.985**		
High	Yes			
High	Yes			
High	No			
High	Yes			
High	No			

Humidity	Play Golf	Entropy	Prob(No)	Prob(Yes)
Normal	Yes	6 Yes,1 No	0.142857143	0.857142857
Normal	Yes	**0.591**		
Normal	Yes			
Normal	Yes			
Normal	Yes			
Normal	No			
Normal	Yes			

Note:

The entropy of the **Information Gain for Humidity**-

0.94– (7/14 * 0.985 + 7/14 * 0.591) = 0.152

Information Gain for Windy

Windy	Play Golf
FALSE	No
FALSE	No
FALSE	Yes
FALSE	Yes
FALSE	Yes
FALSE	Yes
FALSE	Yes
FALSE	Yes
TRUE	Yes
TRUE	Yes
TRUE	No
TRUE	No
TRUE	No
TRUE	Yes

Windy	Play Golf	Entropy	Prob(No)	Prob(Yes)
FALSE	No	2No,6 Yes	0.25	0.75
FALSE	No	**0.81**		
FALSE	Yes			
FALSE	Yes			
FALSE	Yes			
FALSE	Yes			
FALSE	Yes			
FALSE	Yes			

Windy	Play Golf	Entropy	Prob(No)	Prob(Yes)
TRUE	Yes	3 Yes,3 No	0.5	0.5
TRUE	Yes	**1**		
TRUE	No			
TRUE	No			
TRUE	No			
TRUE	Yes			

Note:

The entropy of the **Information Gain for Windy**-

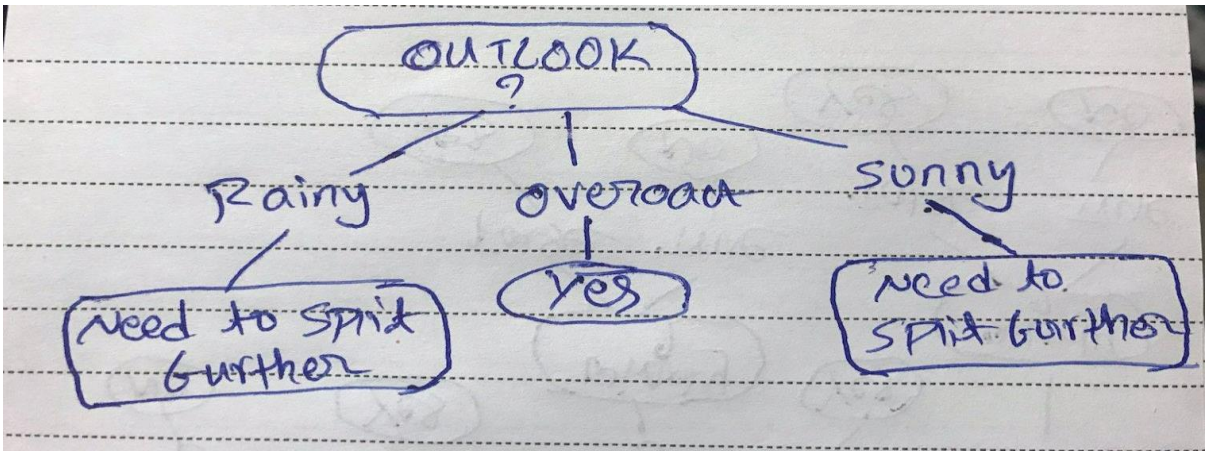
0.94– (8/14 * 0.81 + 6/14 * 1) = 0.04857142857

Table below summarizes the information gain for all four Predictors.

Results of 1st Iteration -

Gain	Outlook	Temp	Humidity	Windy
1st Iteration	**0.25**	**0.029**	**0.152**	**0.049**

- **Outlook** is selected as the root because it has the highest information gain.



Note:

- Since **Overoact** Outlook have been associated with pure target I.e **Yes**, we do not need these data any longer. Since Predictor Outlook has been used in the decision tree, we can remove the Predictor and focus only on the remaining Three Predictors : **Temp, Humidity, Windy.**
- For **second iteration,** our data table is come from the Rainy & Sunny Outlook because it is not associated with pure target. Since, Both Rainy and Sunny have the same information gain value, then pick the Predictor which comes first (as found from left to right) in the dataset. then pick the Predictor which comes first (as found from left to right) in the dataset I.e **Rainy**.

➢ 2nd Iteration

Predictors				**Target**
Outlook	**Temp**	**Humidity**	**Windy**	**Play Golf**
Rainy	Hot	High	FALSE	No
	Hot	High	TRUE	No
	Mild	High	FALSE	No
	Cool	Normal	FALSE	Yes
	Mild	Normal	TRUE	Yes

3N,2Y

Prob(No) - 3/5 ==> 0.6

Prob(Yes) - 2/5 ==> 0.4

Entropy : -3/5 * log2(3/5) - 2/5 * log2(2/5) ==> 0.97

Information Gain for Temperature

Temp	Play Golf
Hot	No
Hot	No
Mild	No
Cool	Yes
Mild	Yes

Temp	Play Golf	Entropy	Prob(No)
Hot	No	2No	1
Hot	No	**0**	

Temp	Play Golf	Entropy	Prob(No)	Prob(Yes)
Mild	No	1No,1Yes	0.5	0.5
Mild	Yes	**1**		

Temp	Play Golf	Entropy	Prob(Yes)
Cool	Yes	1 Yes	1
		0	

Note:

The entropy of the **Information Gain for Temperature** -

0.97– (2/5 * 0 + 2/5 * 1 + 1/5 * 0) = 0.57

Information Gain for Humidity

Humidity	Play Golf
High	No
High	No
High	No
Normal	Yes
Normal	Yes

Humidity	Play Golf	Entropy	Prob(No)
High	No	3 No	1
High	No	**0**	
High	No		

Humidity	Play Golf	Entropy	Prob(Yes)
Normal	Yes	2Yes	1
Normal	Yes	**0**	

Note:

The entropy of the **Information Gain for Humidity**-

0.97– (3/5 * 0 + 2/5 * 0) = 0.97

Information Gain for Windy

Windy	Play Golf
FALSE	No
TRUE	No
FALSE	No
FALSE	Yes
TRUE	Yes

Windy	Play Golf	Entropy	Prob(No)	Prob(Yes)
FALSE	No	2No,1Yes	0.666666667	0.333333333
FALSE	No	**0.9183**		
FALSE	Yes			

Windy	Play Golf	Entropy	Prob(No)	Prob(Yes)
TRUE	No	1No,1Yes	0.5	0.5
TRUE	Yes	**1**		

Note:

The entropy of the **Information Gain for Windy**-

0.97– (3/5 * 0.9183+ 2/5 * 1) = 0.01902

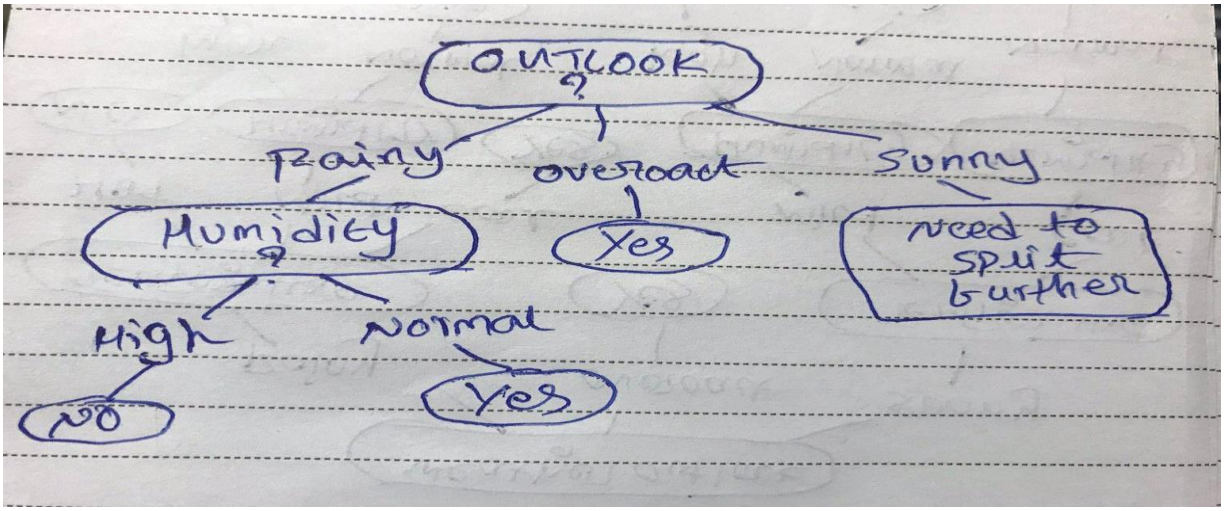
Table below summarizes the information gain for all three Predictors.

Results of 2nd Iteration -

Gain	Temp	Humidity	Windy
2nd Iteration	**0.57**	**0.97**	**0.01902**

Note:

The maximum gain is obtained for the optimum Predictor is Humidity. **This Humidity will be the root of the subtree for Rainy.**



Note:

- Since, **High Humidity & Normal Humidity for Rainy Outlook** have been associated with pure target I.e No & Yes Respectively, we do not need these data any longer.
- For **Third iteration,** our data table is come from the **Sunny** Outlook because it is not associated with pure target.

➢ 3rd Iteration

Predictors				Target
Outlook	**Temp**	**Humidity**	**Windy**	**Play Golf**
Sunny	Mild	High	FALSE	Yes
	Cool	Normal	FALSE	Yes
	Cool	Normal	TRUE	No
	Mild	Normal	FALSE	Yes
	Mild	High	TRUE	No

3Y,2N

Prob(No) - 2/5 ==> 0.4

Prob(Yes) - 3/5 ==> 0.6

Entropy : -3/5 * log2(3/5) - 2/5 * log2(2/5) ==> 0.97

Information Gain for Temperature

Temp	Play Golf
Mild	Yes
Cool	Yes
Cool	No
Mild	Yes
Mild	No

Temp	Play Golf	Entropy	Prob(No)	Prob(Yes)
Mild	Yes	2Yes,1No	0.333333333	0.666666667
Mild	Yes	**0.918**		
Mild	No			

Temp	Play Golf	Entropy	Prob(No)	Prob(Yes)
Cool	Yes	1Yes,1No	0.5	0.5
Cool	No	**1**		

Note:

The entropy of the **Information Gain for Temperature** -

0.97– (3/5 * 0.918 + 2/5 * 1) = 0.0192

Information Gain for Humidity

Humidity	Play Golf
High	Yes
Normal	Yes
Normal	No
Normal	Yes
High	No

Humidity	Play Golf	Entropy	Prob(No)	Prob(Yes)
High	Yes	1Yes,1No	0.5	0.5
High	No	**1**		

Humidity	Play Golf	Entropy	Prob(No)	Prob(Yes)
Normal	Yes	2Yes,1No	0.333333333	0.666666667
Normal	No	**0.918**		
Normal	Yes			

Note:

The entropy of the **Information Gain for Humidity** -

0.97– (3/5 * 0.918 + 2/5 * 1) = 0.0192

Information Gain for Windy

Windy	Play Golf
FALSE	Yes
FALSE	Yes
TRUE	No
FALSE	Yes
TRUE	No

Windy	Play Golf	Entropy	Prob(Yes)
FALSE	Yes	3Yes	1
FALSE	Yes	**0**	
FALSE	Yes		

Windy	Play Golf	Entropy	Prob(No)
TRUE	No	2No	1
TRUE	No	**0**	

Note:

The entropy of the **Information Gain for Windy** -

0.97– (3/5 * 0 + 2/5 * 0) = 0.97

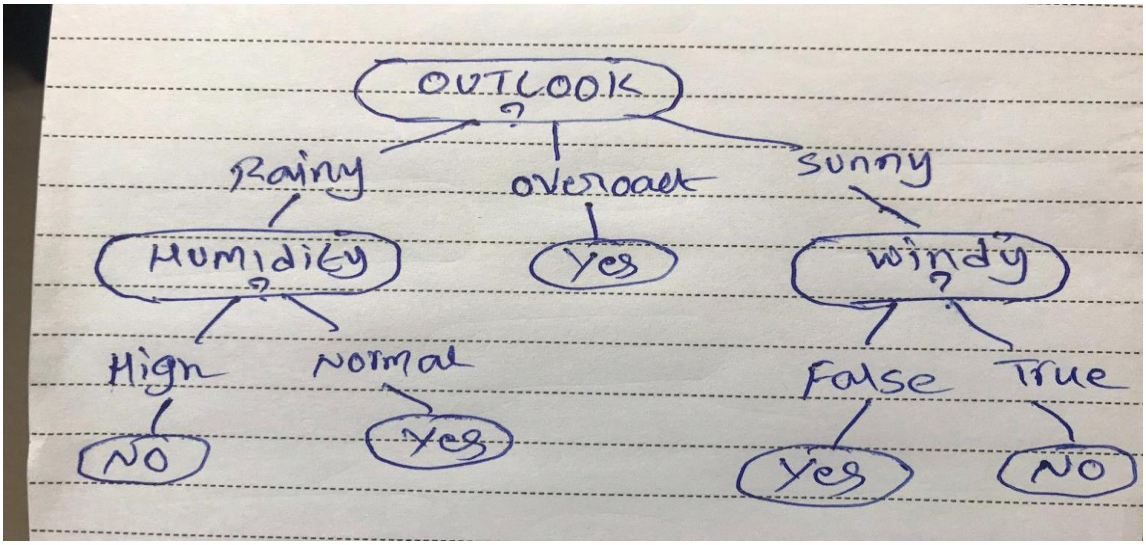
Table below summarizes the information gain for all three Predictors.

Results of 3rd Iteration -

Gain	Temp	Humidity	Windy
3rd Iteration	0.0192	0.0192	0.97

Note:

The maximum gain is obtained for the optimum Predictor is Windy. **This Windy will be the root of the subtree for Sunny.**



Note:

- Since, False Windy & True Windy have been associated with pure target I.e Yes & No Respectively. **we can update our decision tree into the final version.**